meet Archie

The Internet File Finder

The long, hard (ftp) search is over. Let archie, a program that searches Internet files, do all the work for you.

This is the first installment of a four-part series.

By Kevin M. Savetz

Looking for something on the Internet? If you use File Transfer Protocol (ftp), you already know that you can download the latest shareware and public domain games, utilities and applications. Perhaps you already know about the vast archives of electronic books and picture, sound and music files, too.

But if you've used ftp, you've probably also had trouble finding what you're looking for. For one thing, you have to know exactly what file you want and where on the Internet to look for it. And there are so many ftp sites whose contents are usually indexed in a manner that would make a librarian break down and cry. The bottom line: It's not particularly easy to find things using ftp.

That's why you should acquaint yourself with archie, a program that searches for files at anonymous ftp sites. All you need to know is a file's name (or part of it), and archie will give you the exact filename and point you to the file's ftp site and directory. If you've got a hankering to see the Mona Lisa, for example, but don't have the time or funds to trek to Italy, you can ask archie to find a digitized version of Mona.

At last count, archie's database consisted of more than 2.1 million file names (totaling some 170 gigabytes of information) on well over 1,200 anonymous ftp sites. When you start an archie search, the program doesn't try to log into all 1,200 sites at once to look for your file. Instead, archie does its magic by searching from a pre-compiled list of files (a database that weighs in at about 250 megabytes). About once a month, the archie server trolls the Internet's ftp sites, searching for new file listings. So when you do *your* search, the hardest work has already been done.

Accessing Archie Using Telnet

To telnet to a public archie client, select the nearest site from the list below. For a current list of *all* public archie servers worldwide, telnet to a site below and then type servers at the command line.

After telneting, login as archie and, if asked for a password, type archie again. Once you're connected, you can do a basic search by typing find followed by a program name.

Telnet address	IP address	Location
archie.unl.edu	129.93.1.14	Nebraska
archie.internic.net	198.49.45.10	New Jersey
archie.rutgers.edu	128.6.18.15	New Jersey
archie.ans.net	147.225.1.10	New York
archie.sura.net	128.167.254.179	Maryland
archie.univie.ac.at	131.130.1.23	Austria
archie.cs.mcgill.ca	132.206.51.250	Canada
archie.uqam.ca	132.208.250.10	Canada
archie.funet.fi	128.214.6.100	Finland
archie.th-darmstadt.de	130.83.22.60	Germany
archie.unipi.it	131.114.21.10	Italy
archie.wide.ad.jp	133.4.3.6	Japan
archie.sogang.ac.kr	163.239.1.11	Korea
archie.nz	130.195.9.4	New Zealand
archie.rediris.es	130.206.1.2	Spain
archie.switch.ch	130.59.1.40	Switzerland
archie.ncu.edu.tw	140.115.19.24	Taiwan
archie.doc.ic.ac.uk	146.169.11.3	UK

Archie Isn't Perfect

Like most things in life, archie is not perfect. Although it does an admirable job searching for files based on their filenames, it won't search for programs according to what they do. For example, you can't ask for a list of pictures of the Mona Lisa in GIF format or for all *Tetris* programs for the Macintosh.

Archie doesn't know how to ignore old versions of software either. If you want to find only the latest, greatest, feature-laden version of a particular game, for example, you're out of luck. archie will happily list all occurrences of the game, though, if you don't mind sifting through them yourself.

Archie also doesn't have a function for excluding files older than a certain age. Wouldn't it be nice if you could limit a search to files with creation dates less than 120 days ago? Well, you can't, although you *can* ask archie to show the most recently modified files first.

Finally, although archie offers a variety of powerful searching methods, it is still rather unforgiving. If you search for a file called "mona-lisa," for example, but the file is stored in the archives as "mona," you won't find what you're looking for.

Using Archie

There are several ways to access archie—all you need is an archie client, the frontend to the database itself. If you have a shell account, your service provider may have archie already installed on your system. To find out, type **archie** at the command-line prompt. If you see a program rather than "command not found," the archie client is on your system. You can then type **man archie** from the command line for a full list of archie commands available.

If the program hasn't been installed on your system, you have a few options. You can access archie by telneting to one of the many public archie servers scattered around the globe. Or, if you don't have access to telnet, you can send your archie search requests via e-mail. Although e-mail can be used by just about everyone on the 'Net and query results will be dutifully mailed back to you, electronic mail searches seem to take a lot longer than interactive ones.

To use archie via e-mail, send a message to **archie@archie.cs.mcgill.ca** with your search request in the body of the message. In fact, you should be able to send e-mail to any of the archie servers listed in the sidebar. For instance, you can also send e-mail to **archie@archie.rutgers.edu**.

No matter which interface you use, archie uses one basic set of commands.

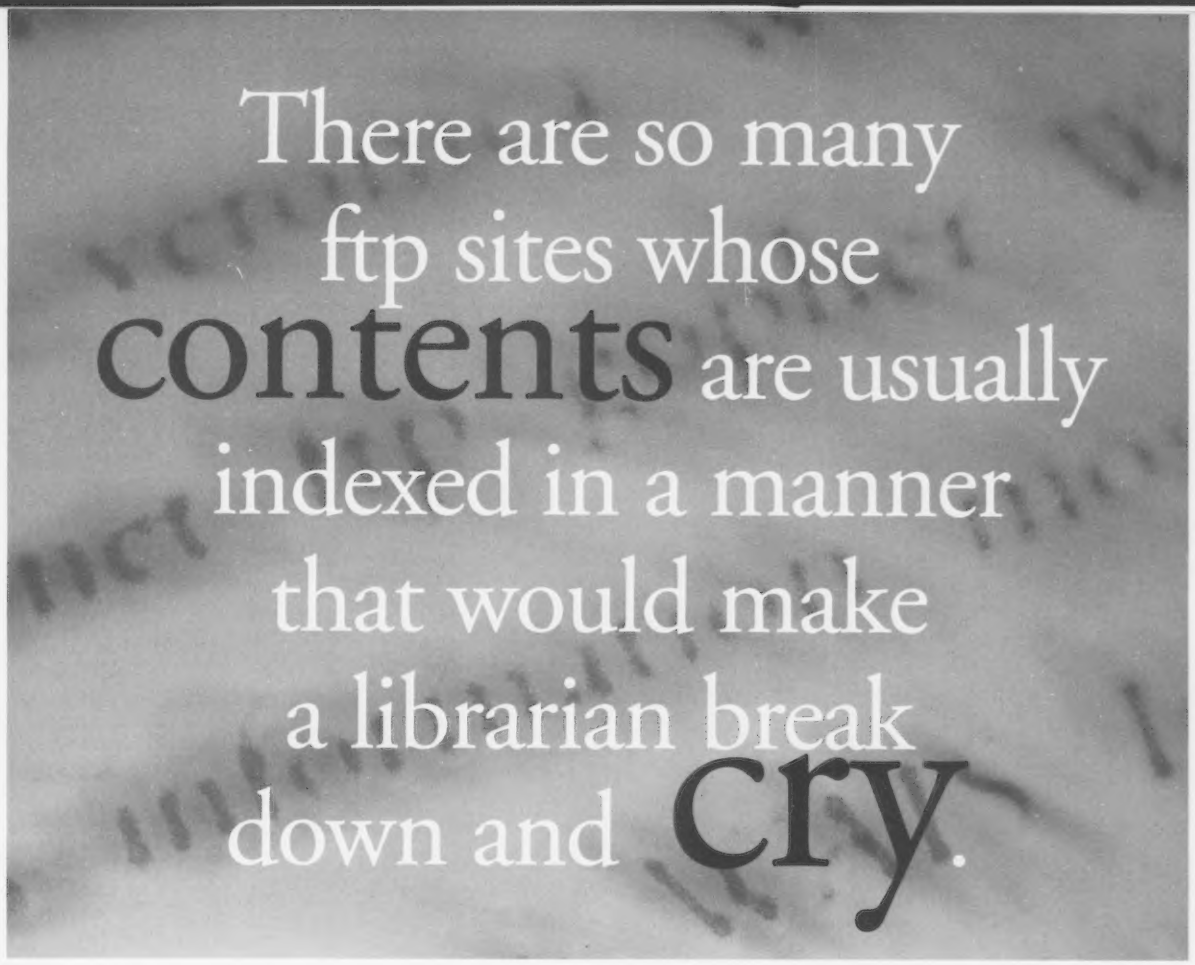
Searching

Find is the command that does archie's real work—searching the database for the file you're looking for. It works simply enough: Just type **find *searchterm***, and archie will show you the files that match your query.

Archie also includes a series of commands enabling you to change how the program will match file names when searching the database. The command, appropriately enough, is called **set search**.

Set search sub, the default searching mode, performs a case-insensitive substring search. A match occurs if any part of the filename contains the item in your find command, without regard to case. For instance, searching for "mon" will hit both "mona-lisa" and "Jokes_From_MONTY_PYTHON."

Set search exact, the fastest search mode, will hit only those files that exactly match your search word—no questions asked. This option is case-sensitive, so searching for "Mona.gif" will hit only on "Mona.gif."



Sorting

To help you find what you're looking for with the greatest of ease, archie can sort its hits before displaying them. You choose the sorting method—via a **set sortby** command—and archie will present its information in the manner that is most useful to you.

Set sortby none doesn't sort the hits at all—archie presents them in the order that they were found. Oddly enough, set sortby none is the default, although it is clearly the least useful way for archie to sort. Not doing a sort is, however, faster than doing one.

Set sortby filename sorts files alphabetically by name. If you want them sorted backwards alphabetically, use the set sortby rfilename command.

Set sortby hostname sorts alphabetically by the ftp site's host name. Again, to sort backwards, the command is **set sortby rhostname**.

Set sortby size focuses on file size—listing the largest files first. Try **set sortby rsize** to sort from smallest to largest.

Set sortby time sorts by file modification date, with the most recently changed files first. You can see the oldest matching files at the top of the list with the **set sortby rtime** command.

Finally, **set maxhits** instructs archie to show you only up to a specified number of matches. The program ordinarily searches all sites when you do a search. If you find that you are inundated with information after performing a few searches, you may want to take advantage of the fact that you can limit the search scope. By default, archie will show you the first 100 matches. If you want to see more, you can type a higher number, up to a maximum of 1,000. On the other hand, you can speed the search by setting maxhits to as low as one.

For More Information

As you can see, archie isn't difficult to use. In fact, you don't need to know any commands other than **find**, but having at least a cursory knowledge of the **set search** and **set sortby** commands can make your archie searches more efficient. Give archie a try the next time you're perusing the endless Internet archives with something in mind — think of it as the Internet's filing cabinet.

If you have any questions about archie, send e-mail to the developers at **info@bunyip.com**. Comments, suggestions and bug reports should be sent to **archie-l@cs.mcgill.ca**.

There's also a mailing list for discussion of the archie software. To subscribe, send your request to **archie-people-request@bunyip.com**.

Next month: Veronica OA

---

*Kevin Savetz (**savetz@northcoast.com**) is a technology writer based in Humboldt County, California. He is author of **Your Internet Consultant: The FAQs of Life Online** (Sams Publishing) and publishes the **Internet Services FAQ** and the **Unofficial Internet Book List**.*